

Social Media Data in Tourism Planning: Analysing Tourists' Satisfaction in Space and Time

Roberta Floris, Michele Campagna

(Roberta Floris, University of Cagliari, Via Marengo 2, 09123 Cagliari, Italy, roberta.floris@unica.it)
(Michele Campagna, University of Cagliari, Via Marengo 2, 09123 Cagliari, Italy, campagna@unica.it)

1 ABSTRACT

Social media are playing an increasingly important role as information resource in tourism both for customers (i.e. the tourists), who gather trustworthy information supporting the choice of destinations and services from peers, and for businesses, which can use the same information for improving their marketing strategies. The use of social media data can also offer new opportunities for decision- support in tourism planning. With improved understanding of the motivations of tourists and tailoring tourism service supply, decision making can be facilitated by emphasizing the strengths of tourist destinations for past and potential visitors. However, this kind of information about tourists perceptions and opinions is not always properly analysed by planners. Understanding the user satisfaction, which depend on factors related to both the location and the services that the local industry propose, may offers valuable information in tourism planning at regional and local level.

In the light of the above premises, the goal of the study presented in this paper is to propose an integrated approach to investigate the relationships between tourists satisfaction, destination resources and tourism industry for supporting design and decision-making in regional tourism planning. The methodology developed in the study includes data collection from popular tourism social media platforms (i.e. Booking.com and TripAdvisor.com), and their integration with territorial and tourism data. Spatial and statistical analysis techniques are then applied to elicit insights from tourists perceptions on success factors which may be used in decision-making and planning support. The case study demonstrates the value of social media data and computational social science techniques in tourism planning. The paper concludes with a critical discussion on the potential of using such an approach in more general urban and regional planning setting.

2 INTRODUCTION

This study focuses on tourism phenomenon, analysing relationship between demand, industry and location identified as fundamental variables. The research aims to study tourist preferences on destination and tourism industry services as represented by review judgments collected by two major tourism social networks, namely TripAdvisor.com and Booking.com. The investigations are carried on exploring the potential of publically volunteered comments, for providing useful knowledge about people preferences in space and time. For research purposes, a traditional method for collecting information about such preferences, performed via ad-hoc surveys can be expensive and time consuming. For this reason this work presents an alternative approach, by which tourist preferences for location and services are discovered by processing and analysing publically available social media data. The paper explores three questions related to tourists preferences:

- (1) Which are the most popular destinations?
- (2) Why people chose those destinations?
- (3) What attracts tourists attention and what do they appreciate/disregard?

The underlying assumption is that this kind of study and underlying methods and tools can be used as successfully in urban and regional planning as in tourist planning for in both cases they contribute to take into account a pluralist customer- (or citizens) –oriented view on strategic development issues. From the methodological perspective, the central challenge in answering the questions above is to manage the big amount of available data to discover useful knowledge.

The method builds on a set of spatial analysis and statistics techniques, useful in describing and visualizing the spatial distribution and detecting patterns and hot-spots. In addition textual analytics techniques (Campagna et Al, 2013; Campagna, forthcoming), have been applied in order to discover the knowledge enclosed in the huge amount of qualitative social media comments. The findings provide insights into the Sardinia tourism industry which could aid in the development of new planning approaches. They also offer a benchmark for future comparative trend analysis and directions for tourism policy design. After examining

the past studies on the travel consumers online social networks and the most popular web sites, the paper focuses on the destination choices and judgments represented in numerical and linguistic terms. Then, the research methodology and the early results of the case study are summarized and briefly discussed. The paper concludes with the summary of findings and future steps for extending the presented work.

3 DEVELOPMENTS IN TOURISM SOCIAL MEDIA

In the last decade the degree of interactivity established by the Web 2.0 paradigm enhanced the role of Internet as information source, with a secondary role as opinion source (Grabner et Al, 2012). Tourism is one of the sectors which exploited the advantage of the advances in ICT and in the development of online communities. As a matter of facts, on the supply side the tourism marketing (i.e. the way to promote tourism industry, the different destinations, or the holiday packages) have been totally transformed (Dippelreiter, 2010). It is no surprise to observe that travel and tourism related topics are among the most popular topics in on-line social networks. (Baggio et Al, 2008). Likewise, on the demand side, the travellers use the Internet to obtain tourist information, to share their experiences, to establish relationships with people from various destinations or to purchase travel related products.

According to Chung and Buhalis (2008) Internet supports i) the pre-travel phase, where all the search and bookings can take place online; ii) the in-travel phase, through interactive forums and/or blogging while on the road; and iii) the post-travel phase where people can share experiences, review hotels and destinations, and post photographs and videos from their trips (Conrady, 2007). TripAdvisor and Booking.com are among the most popular platform of the latter kind, and they play a significant role into the online tourism market. They represent an important marketing channel through which destinations and tourism enterprises can reach and persuade potential visitors (Biassoulis, 2002). In fact travellers opinions and personal experiences based on reconstruction of their trips in turn serve as information for others.

In case location is available all these type of information, as all information derived from social network, could be considered as Volunteer Geographic Information (VGI). The term VGI indicates the avalanche of information which every second is shared on the web by users acting as sensors (Goodchild, 2007). According to Sui and Goodchild (2011), more recently the convergence of GIS and social media granted by interoperability of geo-web tools is further enriching the possibility of sharing the knowledge not only about the Earth surface but also about all the social and cultural phenomena there happening. In facts, as Campagna et Al argue (2013), VGI may include both geographic information collected by groups of people within crowdsourcing initiatives and geo-tagged multimedia collected for personal purposes by the Internet users and publicly shared through archives in the cloud.

4 DESTINATION CHOICES AND TOURISTS' PREFERENCES

Knowledge of consumer psychology is extremely important in determining the success of a destination (Rodriguez del Bosque and San Martin, 2008). In this sense, an exploration of psychological concepts such as attitudes, decision-making processes, emotions, experience and satisfaction is necessary for understanding costumers choices and preferences in tourist destination. According to Kuang Hsu et Al (2009) travel motivation is a dynamic concept; it changes from one person to another and from one destination to another. Cooper (2009) pointed out that one popular typology for understanding travel motivation is the push and pull model by Crompton (1979). The push motivations is useful for explaining the desire for travel while the pull motivations explains the actual choice of destination. The Crompton model identifies seven socio-psychological (push) motivations (escape, self-exploration, relaxation, prestige, regression, kinship-enhancement, and social interaction) and two cultural (pull) motivations (novelty and education). According to Crouch et Al (2004), consumers judgments depend basically on the strength of their beliefs or expectations about the quality of various features or attributes associated with services. Personal preferences, like motivations, may be both intrinsic, reflecting individual likes and dislikes, and extrinsic, or socially conditioned. Each opinion strictly depends on tourists direct past experiences with other services of analogous nature Kuang Hsu et Al (2009). Benitez et Al (2007) argue that the "quality that consumers perceive in a service is a function of the magnitude and direction of the gap between expected service and perceived service". Judgments expressed by numbers are easy to interpret, but linguistic information is more difficult to measure through a mathematical function. Linguistic information characterizes subjective knowledge and are characterized by uncertainty, imprecision and ambiguity (Benitez et Al, 2007).



In the light of the above premises the next section reports on the analysis from the spatial perspective of visitors perceptions of tourism businesses in Sardinia. This approach requires to identify regional patterns and key tourism areas, considering tourists preferences, domain location and their connection. The analysis was developed thanks to the interpretation of Social Media Geographic Information (SMGI) with authoritative geographic information (A-GI) in GIS environment.

5 MULTIDIMENSIONAL ANALYSES OF TOURISM SOCIAL MEDIA INFORMATION

In order to understand the tourist preference dynamics in Sardinia, as expressed by SMGI, a two scales approach was adopted. Firstly analyses at the regional scale were carried on to describe tourists preferences spatial patterns and to identify location of interest; the latter may include clusters of positive or negative preferences, or individual spots of interest. Then, at the local level (i.e. within the single cluster or spot of interest) further analysis were carried on aiming at understanding the possible reasons beneath the patterns and singularities with the assumption that they may help in explaining success or failure factors with regards to destination and services features. Both at the regional and the local levels, an investigation method was adopted including descriptive spatial analysis and spatial statistics coupled with explanatory SMGI analyses, including Spatial-Temporal Textual analysis, which can be defined as the textual analysis constrained by space and time boundaries (STTx; Campagna, forthcoming).

Operationally the study was carried on according to the following workflow: i) Data collection and geocoding: data were extracted by Booking.com and TripAdvisor.com, geocoded and integrated in a geodatabase for analyses; ii) Regional preferences dynamics analysis: data were analysed for all the region at the municipal unit of analysis with spatial analysis, spatial statistics, and STTx in order to detect clusters and hot/cold-spot; iii) Local preferences dynamics analysis: data were integrated with authoritative data from the regional Spatial Data Infrastructure and other official open data sources in order to find explanatory hints on the preference dynamics and to get deeper insights on the relationships among tourist preferences, local territorial features and quality of industry services in selected destinations. The last two steps were carried on iteratively on the relevant clusters and spots as in the examples reported in the remainder of this section.

5.1 Data collection and geocoding

In the first step, of the study a database was created extracting data from TripAdvisor.com and Booking.com in the period between May 2012 and May 2013. Through these applications customers can book, rank and review hotels, flights and restaurants (or Tourism Services, TS). The focus of the portals is to filter content based on rankings that are derived from other users ratings. Thus, rankings are split into several categories, such as value/price, rooms, location, cleanliness and sleep quality. Available rating categories however are determined by the type of reviewed item. The reviews are enriched by the possibility to add multimedia elements or travel maps of previous trips or to take part in discussion forums.

Thanks to the availability of the location of the services they can thus be considered SMGI (Campagna, forthcoming). Thus, the study required the adoption of a mixed methods approach, in which quantitative and qualitative information were collected in a database for analyses. The quantitative information concerns the score of tourist evaluation criteria, while qualitative information includes customers textual descriptive review. Concerning the quantitative analysis it should be noted that in TripAdvisor.com a rating scale consists of five ordinal values (or stars), ranging from 'terrible' to 'excellent'. A separate mandatory overall rating summarizes the total customer satisfaction. In Booking.com a rating scale consist of numerical integer ordinal values, ranging from 1 to 10 (i.e. the higher the better). Beside quantitative assessment, in both platforms, a text box records qualitative natural language reviews. The title is a concise short text formulation of the assessment, while the comment is a long text field. After the data collection, a geodatabase was created including 992 Tourism Lodging Service (TLS) records extracted from TripAdvisor.com and Booking.com (the same TLS from both websites). The records provide TLS name, category, location, and related quantitative score. It should be noted that the TLS category includes not only hotels, but also other types of accommodation such as resort, bed&breakfast or agritourism. In order to analyse spatially the location of the tourism business patterns in Sardinia, geocoding was performed on the extracted addresses, providing the exact location of the tourism operators.

TLS were divided into 5 main categories: agritourisms (6%), bed and breakfast (15.7%), hotels (42%), private accommodations (29%), residences and resorts (7.3% of total numbers of operators). the analysis by

provinces revealed that the 3 provinces have emerged as important tourist destinations in the tourists perception: Olbia-Tempio (27,8%) Sassari (24%) and Cagliari (20,6%). Other 4 provinces (Nuoro 8,6%, Oristano 7,3%, Ogliastra 5%; Carbonia-Iglesias 4%) are well represented by tourism businesses; whilst the province of Medio Campidano is only represented by the 3% tourism businesses. In addition, analysis of the significance of tourist appreciation in the coast and in inner areas in Sardinia revealed that 92% of tourism reviews sample concern LTS (917) in the coastal areas, while only less than 8% of popular tourism businesses are found inland. Nevertheless, Nuoro and Medio Campidano Provinces together provide notable inland popular TLS with almost the 13% of the total number of reviewed tourism businesses. This may mean that tourists visit these areas to discover a less popular side of the island, which is characterized by its nature, cultural heritage and traditions.

5.2 Spatial analysis of ourism preferences

After the preliminary descriptive analyses of the preferences dataset, the second step of the methodology is the application of spatial analyses of tourism preferences to explore spatial patterns of positive visitors judgments at the regional level. The application of spatial analytical techniques allows the exploration of the spatial dynamics of visitors perception and their relationships with different variables. For each TLS the database includes a score record, which is the average of six main attributes: 1) Location, which is related to the geographic position of the structure; 2) Services, referring to all transport facilities, shopping areas, bars and restaurants; 3) Price/quality ratio, referring to structure clearness, staff kindness and all type of comforts offered by the operators; 4) Staff (kindness); 5) Room cleanness (Cleaning); 6) Comfort, referring to all facilities and services that hotels provide to their customers.

It should be noted that the attributes Location and Services are those which explain the territorial features of the destination while the others express the perceived quality of the TLS supply.

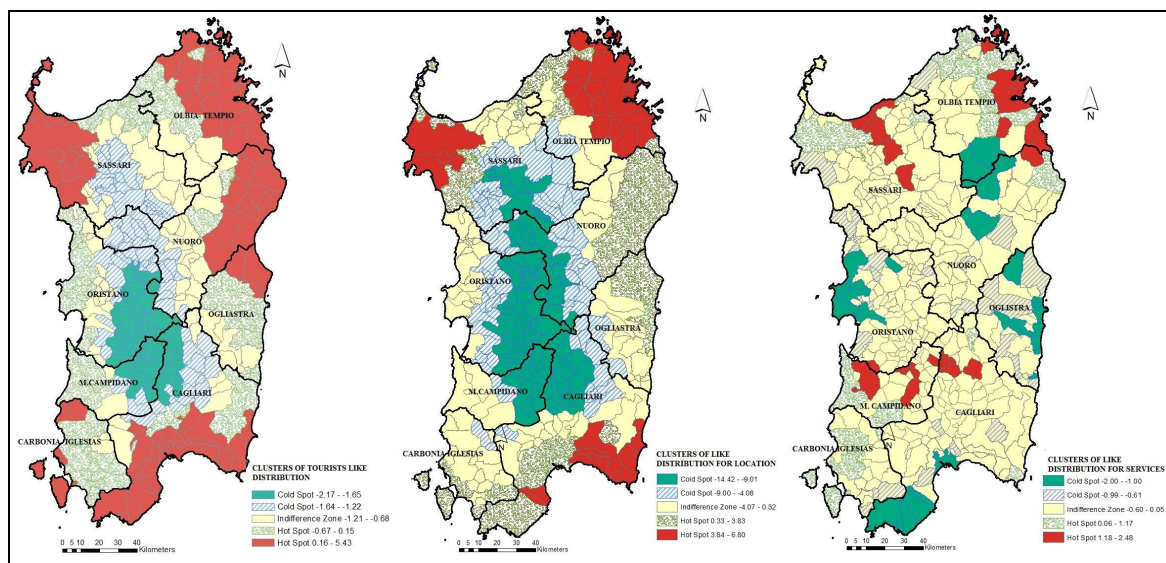


Figure 1 Areas characterized by positive preferences (TPPI): global (left), location (center), and services (right).

Thus the data model allows the investigation of the spatial patterns of both preferences on territorial and tourism industry features at the local level across the all region. The following analyses show some examples, which explains which are globally the favourite destinations and by the two perspectives. The analysis starts by mapping the Tourist Positive Preferences Incidence (TPPI, i.e. the ratio between the positive scores and the TLS by municipality) in Sardinia. Figure 1 shows the distribution of the TPPI (left). The TPPI shows an overall high spatial concentration in the North-East of Sardinia. The Costa Smeralda district appears as the only area where the global tourism preferences fulfil overall visitors expectations. Looking at individual municipalities, the analysis shows that Alghero is the one with the highest TPPI rate. The other two municipalities with a high TPPI are Cagliari and Olbia. Also in Figure 1, two maps show the pattern of TPPI by destination territorial features (centre) and by tourist industry services quality (right).

5.3 SMGI analysis at the local level

After the analysis of the tourism dynamics described the preference patterns at the regional level identifying clusters and spots of successful destinations the methodology adopted for this study shifted to the local scale for further analyses aiming at finding explanatory answers for the phenomena under observation. The shift from the regional to the local scale is also carried on relying on spatial analysis and spatial statistics techniques on an integrated SMGI/A-GI data database. As an example case study for the sake of illustrating the methodology steps, the tourist destination of Alghero have been chosen as the regional analyses demonstrated its highly successful performance. The analyses at the local scale are intended to investigate the success factors within the single destination in order to extract useful hints to be used for further planning in the same or other destinations.

Alghero has been recognized as a best-selling destination from different tourists typologies. Thus, the following questions one should answer were Why tourists interest concentrates in Alghero? and What exactly in the destination does attract the tourists attention? In order to answer these questions, summarising the review by neighbourhoods the analysis of preferences spatial clusters reveals that the historic city centre of Alghero attracted the main attention of the visitors, while the modern residential districts in the outskirts, in this case, represents a cold spot. This kind of research can be supported by the integration of SMGI data with other A-GI on demographic, land use, transport facilities or socio-economic data coming from the regional SDI. In this sense, one interesting research question is whether spatial statistic methods such as regression analysis can be used to understand whether the spatial interest of the participant is influenced by environmental or socio-cultural variables. This is represents the next future step in the extension of the study.

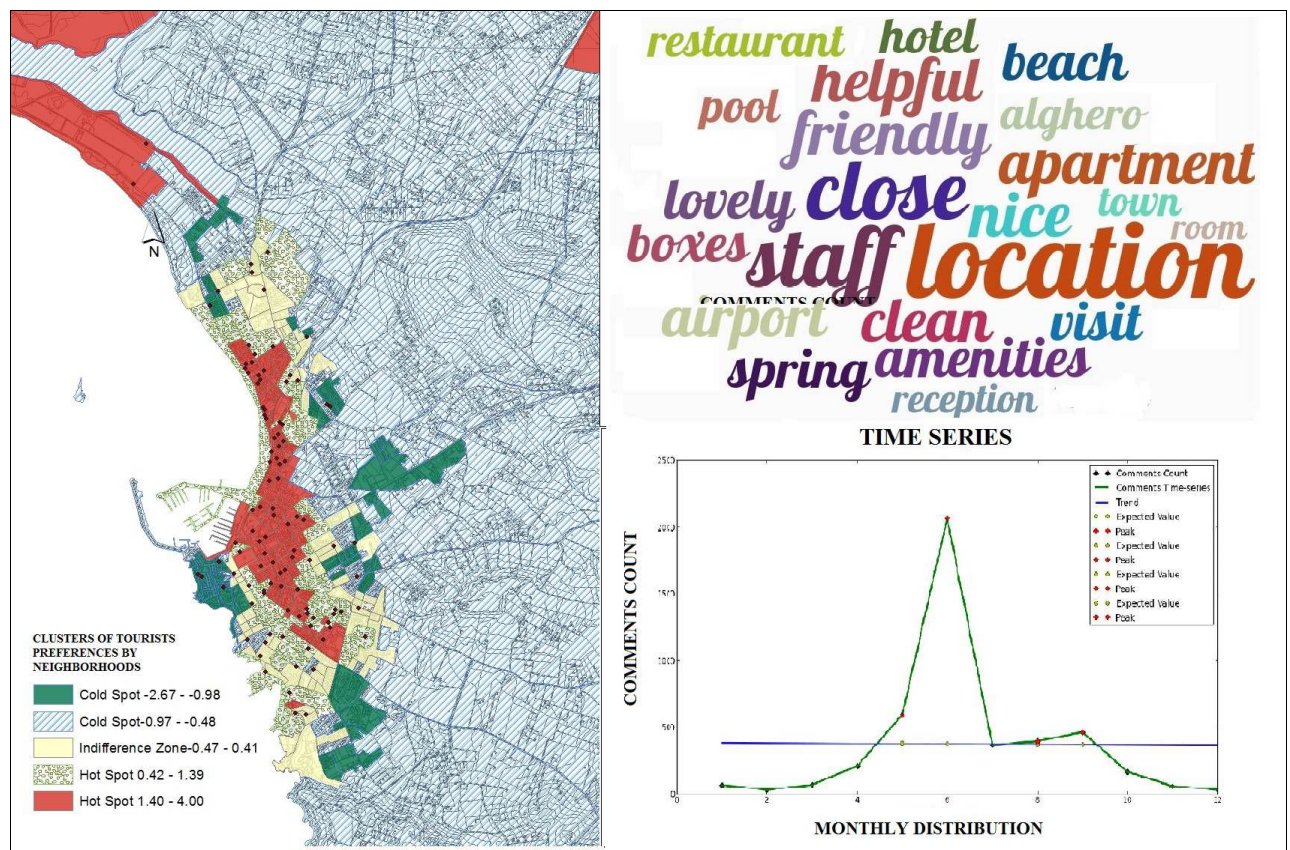


Figure 2 Significant patterns in Alghero (left), tag cloud for hot spots affected areas (top right) and time series (bottom right)

Another effective way to investigate the why tourist demonstrate to prefer certain areas or destinations rather than other may be given by the STTx analysis on their reviews. We want understand not only where but also what people think, analysing the reviews content. The analysis was carried out using the tourist comment database, which collected data regarding tourist user origin, language used, time comment and the textual judgment (positive and negative). For the overall Sardinia more than 880.000 reviews were extracted in 5 different languages (Italian, Spanish, English, French and Portuguese).

The textual analysis using Tag Cloud led to discover knowledge enclosed into this huge amount of text comments. The Tag cloud is essentially a visual representation of labels (tag) or keywords contained in different word strings. Generally, this representation is presented with a weighted word list in alphabetical order, where the larger font is attributed to the most repeated words (Kaser, 2007). In this example (Figure 2, left part), the most popular 20 words were extracted. The analysis reveals that the majority of the words in the posts refer to spatial or physical aspects of Alghero, such as location, beach, town, and city centre. Other frequent words are related to tourism structures, such as hotel, stuff, room and pool. Outcomes of the textual analysis also indicate higher levels of satisfaction with location, facilities and services.

According to the results, the main reason for tourists to visit Alghero seems to be related to both its natural attractions, which include natural sites, such as beaches, and the presence of a unique cultural heritage. These facts generate a positive tourism location image, which are the most influential psychological factors at play when tourists decide where to travel. Results also indicate a high level of satisfaction with the destination leisure sites, such as typical restaurants and typical food. In addition, results expressed a high level of satisfaction with the supply of accommodation, the cleanliness of structures and the kindness of the employed staff.

In order to identify other significant patterns, the outcomes of the spatial distribution of tourist comments analysis, have been further investigated using the temporal analysis. Figure 2 (bottom right) reports the tourists activity degree by month. The activity is higher in the beginning of the summer season, especially in June and July. This is a very interesting time behaviour to analyse and represents an additional step of the study.

6 DISCUSSIONS AND CONCLUSIONS

The results presented in this paper rely on the first outcome of an exploratory study to get insight on what kind of analysis may be carried on in order to extract from Social Media and Geographic Information meaningful knowledge relevant for planning and decision-making. The case study falls in the domain of tourist planning which is closely related to urban and regional planning. In fact, as the results of this study demonstrate, the success of tourist destination is closely dependent not only by the quality of the tourist industry offer but also by the territorial setting of the destinations, including the natural, cultural and the physical character of the places, as well as infrastructure and services. Further analyses are currently ongoing to extend this early framework and to earn deeper insights on the one hand on the functioning of tourism preference dynamics, and on the other hand, from the methodology perspective, on the formalization of a novel and robust integrated A-GI/SMGI analytics.

Still, this study gives empirical contributions to the evaluation of social media data using spatial analysis tools in tourism literature. The first one is related to the use of exploratory spatial analysis as a method to visualize and interpret visitors perception based scores. The literature on tourism services distribution highlights several issues and debates, but often the spatial dimensions of visitors subjective perception was omitted so far. In addition, the measure of this spatial dimension and its representation may open new opportunities for planners as well as new research challenges, in order to use authoritative and social media GI for a pluralist and customer-oriented policy-making in tourism planning. Many of the assumptions and findings can be anyway applied to the more general field of urban and regional analysis, design and planning.

Additional analyses are currently under development aiming at understating possible the integrations of SMGI with a more complex territorial model relying on further official spatial data resources on demographic, land use, transport facilities or socio-economic data coming from regional SDI. In this sense, one interesting research question which will be tested is whether spatial statistic methods such as spatial regression analysis can be used to investigate quantitatively how the spatial interest of the participant is influenced by environmental or socio-cultural variables.

7 REFERENCES

- BAGGIO R., Costa C., Miguens J.: Social media and tourism destinations: Tripadvisor case study. In *Advances in tourism research*, Vol 26, Issue 28, 2008.
- BENITEZ J., Martin J., Roman C.: Using fuzzy number for measuring quality of service in the hotel industry. In *Tourism management*. Vol 26, pp 544-555, 2007.



- BRIASSOULIS H. : Sustainable tourism and the question of the commons. In *Annals of Tourism Research*. Vol 29, Issue 4, pp 1065–1085. 2002.
- CAMPAGNA M. et All. : Place I care! Crowdsourcing planning information. In AESOP-ACSP Joint Congress. Dublin, 2013.
- CHUNG J. Y., Buhalis D.: Web 2.0: A study of online travel community. In *Information and communication technologies in tourism*, pp 70-81. 2008.
- CONRADY R. : Travel technology in the era of Web 2.0. In *Trand and issues in Global Tuorism*, pp 165-184. Berlin, 2007.
- COOPER C., Fletcher J., Fyall A., Gilbert D., Wahill S. : *Tourism: Principles and Practice*. Harlow, 2008.
- CROMPTON J.. :Motivations for pleasure travel. *Annual of Tourism Research*..Vol 6, pp 408–424. 1979,
- CROUCH G, Perdue R., Timmermans H., Uysal M.: *Consumer Psychology of Tourism*. In *Hospitality and Leisure*, Vol 3. London, 2004.
- DELLAERT B., Dick F. Ettema L., Christer L.: Multi-faceted tourist travel decisions: a constraint-based conceptual framework to describe tourists sequential choices of travel components. *Tourism Management*. Vol 19, Issue 4, pp 313-320. 1998.
- DIPPELREITER B. et Al.: Online tourism communities on the path to Web 2.0: an evaluation. In *Information technology & tourism*. Vol 10, Issue 4. 2008.
- FOTHERINGHAM S.A, Brunsdon C., Charlton M.: *Geogrgraphically weighted regression: the analysis of spatially varying relationships*. 2002.
- GOODCHILD M. Citizen as Voluntary sensors: spatial data infrastructure in the World of Web 2.0. In *International Journal of Spatial Data Infrastructures*. Vol 2, pp 24-32. 2007.
- GRÄBNER D.: Classification of customer reviews based on sentiment analysis. *Information and Communication Technologies in Tourism*, pp 460-470. Springer,Vienna. 2012..
- LAZER D., Pentland A., Adamic L., Aral S., Barabasi A.L., Brewer D., Christakis N., Contractor N., Fowler J., Gutmann M., Jebara T., King G., Macy M., Roy D., Van Alstyne M.: *Computational Social Science*. Science. Vol 323, pp 721–723. 2009.
- LEWIS R., Booms B. H.: *the marketing aspects of service quality. Emerging perspectives on services marketing*. Pp 99-107. Upah Chicago. 1983.
- KUANG TZU H., Tsai Y., Wu H.: The preference analysis for tourist choice of destination: A case study of Taiwan. *Tourism Management*. Vol. 30, Issue 2, pp 288-297. 2009:
- RODRIGUEZ DEL BOSQUE I., San Martin H., *Tourist satisfaction. A cognitive affective model*. In *Annals of tourism research*., Vol 35, Issue 2, pp 551-573. 2008.
- SUI D., Goodchild M.: The convergence of GIS and social media: challenges for GIScience. In *International Journal of Geographical Information Science*. Vol 25, pp 1737–1748. 2011.

8 ACKNOWLEDGEMENTS

The work presented in this paper was developed by the author within the research project “Efficacia ed efficienza della governance paesaggistica e territoriale in Sardegna: il ruolo della VAS e delle IDT” [Efficacy and efficiency of landscape and environmental management in Sardinia: the role of SEA and of SDI] CUP: J81J11001420007 funded by the Autonomous Region of Sardinia under the Regional Law n° 7/2007 "Promozione della ricerca scientifica e dell'innovazione tecnologica in Sardegna".